



[About](#) [Citation Policy](#) [Donate a Data Set](#)
[Contact](#)

☒ Repository ☐ Web 

[View ALL Data Sets](#)

Abalone Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: Predict the age of abalone from physical measurements



Data Set Characteristics:	Multivariate	Number of Instances:	4177	Area:	Life
Attribute Characteristics:	Categorical, Integer, Real	Number of Attributes:	8	Date Donated	1995-12-01
Associated Tasks:	Classification	Missing Values?	No	Number of Web Hits:	321232

Source:

Data comes from an original (non-machine-learning) study:

Warwick J Nash, Tracy L Sellers, Simon R Talbot, Andrew J Cawthorn and Wes B Ford (1994)

"The Population Biology of Abalone (*Haliotis* species) in Tasmania. I. Blacklip Abalone (*H. rubra*) from the North Coast and Islands of Bass Strait",
 Sea Fisheries Division, Technical Report No. 48 (ISSN 1034-3288)

Original Owners of Database:

Marine Resources Division
 Marine Research Laboratories - Taroona
 Department of Primary Industry and Fisheries, Tasmania
 GPO Box 619F, Hobart, Tasmania 7001, Australia
 (contact: Warwick Nash +61 02 277277, [wnash '@' dpi.tas.gov.au](mailto:wnash@dpi.tas.gov.au))

Donor of Database:

Sam Waugh ([Sam.Waugh '@' cs.utas.edu.au](mailto:Sam.Waugh@cs.utas.edu.au))
 Department of Computer Science, University of Tasmania
 GPO Box 252C, Hobart, Tasmania 7001, Australia

Data Set Information:

Predicting the age of abalone from physical measurements. The age of abalone is determined by cutting the shell through the cone, staining it, and counting the number of rings through a microscope – a boring and time-

consuming task. Other measurements, which are easier to obtain, are used to predict the age. Further information, such as weather patterns and location (hence food availability) may be required to solve the problem.

From the original data examples with missing values were removed (the majority having the predicted value missing), and the ranges of the continuous values have been scaled for use with an ANN (by dividing by 200).

Attribute Information:

Given is the attribute name, attribute type, the measurement unit and a brief description. The number of rings is the value to predict: either as a continuous value or as a classification problem.

Name / Data Type / Measurement Unit / Description

Sex / nominal / -- / M, F, and I (infant)
 Length / continuous / mm / Longest shell measurement
 Diameter / continuous / mm / perpendicular to length
 Height / continuous / mm / with meat in shell
 Whole weight / continuous / grams / whole abalone
 Shucked weight / continuous / grams / weight of meat
 Viscera weight / continuous / grams / gut weight (after bleeding)
 Shell weight / continuous / grams / after being dried
 Rings / integer / -- / +1.5 gives the age in years

The readme file contains attribute statistics.

Relevant Papers:

Sam Waugh (1995) "Extending and benchmarking Cascade-Correlation", PhD thesis, Computer Science Department, University of Tasmania.
[\[Web Link\]](#)

David Clark, Zoltan Schreter, Anthony Adams "A Quantitative Comparison of Dystal and Backpropagation", submitted to the Australian Conference on Neural Networks (ACNN'96).

Papers That Cite This Data Set¹:



Ilhan Uysal and H. Altay Guvenir. [Instance-Based Regression by Partitioning Feature Projections](#). Applied. 2004. [\[View Context\]](#).

Jianbin Tan and David L. Dowe. [MML Inference of Decision Graphs with Multi-way Joins and Dynamic Attributes](#). Australian Conference on Artificial Intelligence. 2003. [\[View Context\]](#).

Edward Snelson and Carl Edward Rasmussen and Zoubin Ghahramani. [Warped Gaussian Processes](#). NIPS. 2003. [\[View Context\]](#).

Anton Schwaighofer and Volker Tresp. [Transductive and Inductive Methods for Approximate Gaussian Process Regression](#). NIPS. 2002. [\[View Context\]](#).

Alexander G. Gray and Bernd Fischer and Johann Schumann and Wray L. Buntine. [Automatic Derivation of](#)

[Statistical Algorithms: The EM Family and Beyond](#). NIPS. 2002. [\[View Context\]](#).

Christopher K I Williams and Carl Edward Rasmussen and Anton Schwaighofer and Volker Tresp. [Observations on the Nystrom Method for Gaussian Process Prediction](#). Division of Informatics Gatsby Computational Neuroscience Unit University of Edinburgh University College London. 2002. [\[View Context\]](#).

Marc Sebban and Richard Nock and Stéphane Lallich. [Stopping Criterion for Boosting-Based Data Reduction Techniques: from Binary to Multiclass Problem](#). Journal of Machine Learning Research, 3. 2002. [\[View Context\]](#).

Shai Fine and Katya Scheinberg. [Incremental Learning and Selective Sampling via Parametric Optimization Framework for SVM](#). NIPS. 2001. [\[View Context\]](#).

Iztok Savnik and Peter A. Flach. [Discovery of multivalued dependencies from relations](#). Intell. Data Anal, 4. 2000. [\[View Context\]](#).

Matthew Mullin and Rahul Sukthankar. [Complete Cross-Validation for Nearest Neighbor Classifiers](#). ICML. 2000. [\[View Context\]](#).

Nir Friedman and Iftach Nachman. [Gaussian Process Networks](#). UAI. 2000. [\[View Context\]](#).

Bernhard Pfahringer and Hilan Bensusan and Christophe G. Giraud-Carrier. [Meta-Learning by Landmarking Various Learning Algorithms](#). ICML. 2000. [\[View Context\]](#).

Kai Ming Ting and Ian H. Witten. [Issues in Stacked Generalization](#). J. Artif. Intell. Res. (JAIR, 10. 1999. [\[View Context\]](#).

Tapio Elomaa and Juho Rousu. [General and Efficient Multisplitting of Numerical Attributes](#). Machine Learning, 36. 1999. [\[View Context\]](#).

Christopher J. Merz. [Using Correspondence Analysis to Combine Classifiers](#). Machine Learning, 36. 1999. [\[View Context\]](#).

Khaled A. Alsabti and Sanjay Ranka and Vineet Singh. [CLOUDS: A Decision Tree Classifier for Large Datasets](#). KDD. 1998. [\[View Context\]](#).

Marko Robnik-Sikonja and Igor Kononenko. [Pruning Regression Trees with MDL](#). ECAI. 1998. [\[View Context\]](#).

Christopher J. Merz. [Combining Classifiers Using Correspondence Analysis](#). NIPS. 1997. [\[View Context\]](#).

Luc Hoegaerts and J. A. K Suykens and J. Vandewalle and Bart De Moor. [Subset Based Least Squares Subspace Regression in RKHS](#). Katholieke Universiteit Leuven Department of Electrical Engineering, ESAT-SCD-SISTA. [\[View Context\]](#).

C. Titus Brown and Harry W. Bullen and Sean P. Kelly and Robert K. Xiao and Steven G. Satterfield and John G. Hagedorn and Judith E. Devaney. [Visualization and Data Mining in an 3D Immersive Environment: Summer Project 2003](#). [\[View Context\]](#).

Rong-En Fan and P. -H Chen and C. -J Lin. [Working Set Selection Using the Second Order Information for Training SVM](#). Department of Computer Science and Information Engineering National Taiwan University. [\[View Context\]](#).

Johannes Furnkranz. [Round Robin Rule Learning](#). Austrian Research Institute for Artificial Intelligence. [\[View Context\]](#).

Christian Borgelt and Rudolf Kruse. [Speeding Up Fuzzy Clustering with Neural Network Techniques](#). Research Group Neural Networks and Fuzzy Systems Dept. of Knowledge Processing and Language Engineering, School of Computer Science Otto-von-Guericke-University of Magdeburg. [\[View Context\]](#).

Miguel Moreira and Alain Hertz and Eddy Mayoraz. [Data binarization by discriminant elimination](#). Proceedings of the ICML-99 Workshop: From Machine Learning to. [\[View Context\]](#).

Johannes Furnkranz. [Pairwise Classification as an Ensemble Technique](#). Austrian Research Institute for Artificial Intelligence. [\[View Context\]](#).

Edward Snelson and Carl Edward Rasmussen and Zoubin Ghahramani. [Draft version; accepted for NIPS*03 Warped Gaussian Processes](#). Gatsby Computational Neuroscience Unit University College London. [\[View Context\]](#).

Sally Jo Cunningham. [Dataset cataloging metadata for machine learning applications and research](#). Department of Computer Science University of Waikato. [\[View Context\]](#).

Bernhard Pfahringer and Hilan Bensusan. [Tell me who can learn you and I can tell you who you are: Landmarking Various Learning Algorithms](#). Austrian Research Institute for Artificial Intelligence. [\[View Context\]](#).

. [Efficiently Updating and Tracking the Dominant Kernel Eigenspace](#). (a) Katholieke Universiteit Leuven Department of Electrical Engineering, ESAT-SCD-SISTA. [\[View Context\]](#).

Citation Request:

Please refer to the Machine Learning Repository's [citation policy](#)

[1] Papers were automatically harvested and associated with this data set, in collaboration with [Rexa.info](#)

Supported By:



In Collaboration With:



[About](#) || [Citation Policy](#) || [Donation Policy](#) || [Contact](#) || [CML](#)